



Forecasting meteotsunamis with neural networks: the case of Ciutadella harbour (Balearic Islands)

Maria-del-Mar Vich¹ · Romualdo Romero¹

Received: 1 December 2019 / Accepted: 29 April 2020 / Published online: 25 May 2020
© Springer Nature B.V. 2020

Abstract

This paper explores the applicability of neural networks (NN) for forecasting meteotsunamis affecting Ciutadella harbour (Menorca, Balearic Islands, Spain). Virtually every year, Ciutadella suffers meteotsunamis with wave heights (crest-to-trough difference in about 5-min interval) around 1 m, and at several episodes in its modern history, the waves have reached 2–4 m. A timely and skilled prediction of these phenomena could significantly help to mitigate the damages inflicted to the port facilities and the moored vessels. Once properly trained, a NN is a computationally cheap forecasting method; the approach could be easily incorporated by civil services which are responsible for issuing warnings and organizing a prompt response. We examine the relevant physical mechanisms that promote meteotsunamis in Ciutadella harbour and choose the input variables of the NN accordingly. Two different NNs are devised and tested: a dry and wet scheme. The difference between schemes resides on the input layer, while the first scheme is exclusively focused on the triggering role of atmospheric gravity waves (governed by temperature and wind profiles across the tropospheric column), the second scheme also incorporates humidity as input information with the purpose of accounting for the occasional influence of moist convection. We train both NNs using the resilient backpropagation with weight backtracking method. Their performance is tested by means of classical verification indexes. We also compare both NN results against the performance of a substantially different prognostic method that relies on a sequence of atmospheric and oceanic numerical simulations (TRAM-rissaga method). The new prediction systems work fairly well in distinguishing rissaga and non-rissaga situations, even though they tend to underestimate the amplitude of the harbour oscillation. Both NN schemes show a skill comparable to that of computationally expensive approaches based on direct numerical simulation of the physical mechanisms. The expected greater versatility of the wet scheme over the dry scheme cannot be clearly proved owing to the limited size of the training database, which lacks a sufficient number of convectively driven rissaga events. The results emphasize the potential of a NN approach and open a clear path to an operational implementation using the whole database for training, avoiding the limitations derived from splitting the available list of events into training and testing subsets.

✉ Maria-del-Mar Vich
mar.vich@uib.es

¹ Grup de Meteorologia, Departament de Física, Universitat de les Illes Balears, Ctra. de Valldemossa, km 7.5, 07122 Palma, Illes Balears, Spain

Keywords Meteotsunamis · Rissaga · Neural networks · Forecasting

1 Introduction

Meteotsunamis are atmospherically generated long ocean waves with significant amplitude in the tsunami frequency band (Vilibić et al. 2016). This phenomenon is observed worldwide, especially in bays and narrow inlets, disrupting the normal operations at the port facilities as the sea level variations they produce can be quite big. Even though atmospherically driven sea level oscillations are weaker than tsunamis, when optimally coupled with open-sea and harbour amplification mechanisms they can be as destructive (Rabinovich 2010). These resonant mechanisms rely on the topography of the sea floor and coastal features, like harbour orientation, width and depth (Monserrat et al. 2006). In this paper, we center our attention to those occurring in Ciutadella harbour in Menorca (Balearic Islands, Spain). Since this is a shallow and narrow harbour, it provides the ideal environment for coastal resonance and, occasionally, extreme meteotsunami events (Ramis and Jansà 1983; Rabinovich and Monserrat 1998). In fact, Ciutadella harbour experiences extreme meteotsunamis every 5–10 years on average. A *rissaga*,¹ the local name for a meteotsunami in Ciutadella, is usually observed during the warm season (May to September) with several ordinary cases every year (wave heights smaller than 100 cm). During an extreme rissaga, the port swings from a dry bed near the closed end of the harbour to a tremendous water flow (e.g. the 400 cm meteotsunami of 15 June 2006, Jansà et al. 2007). This surge floods the adjacent areas and breaks moorings, leaving the vessels adrift, all this in about 5 min. Obviously, timely and skilled predictions of these events are greatly demanded by society and its agents to diminish rissagues' devastating consequences as much as possible.

Meteotsunami phenomena have been widely studied in the international context by many researchers which has led to a great understanding of the physical processes that govern these oscillations (e.g. Ramis and Jansà 1983; Tintoré et al. 1988; Rabinovich and Monserrat 1998; Monserrat et al. 2006; Šepić et al. 2015; Vilibić et al. 2016; Horvath et al. 2019, for the Mediterranean and Black Sea regions). The conceptual model of how a meteotsunami is generated over the Mediterranean and Black Sea regions was assembled and illustrated in Šepić et al. 2015 (see Fig. 4 of that paper). The atmospheric synoptic setting induces a tropospheric ducting of internal gravity waves that produce pressure perturbations of a few hPa-scale amplitude at sea level and with characteristic periods of several minutes. Environments with highly sheared winds in the tropospheric column and combining a very stable layer at low levels with low stability of mid-tropospheric air, are especially conducive to these high-amplitude gravity waves. These pressure variations would produce an equivalent of only a few centimeter-scale perturbation of sea surface height. Thus, some sort of maritime amplification mechanism is needed to produce a meteotsunami (Monserrat et al. 2006). A sequence of mechanisms acting on the long ocean waves as they approach the coast have been proposed: (1) Proudman resonance, the ocean waves speed matches the speed of the forcing atmospheric gravity waves; (2) shelf amplification or shoaling; and (3) harbour resonance, the frequency of incoming long-ocean waves matches harbour eigenperiods.

¹ Rissagues is the plural form.

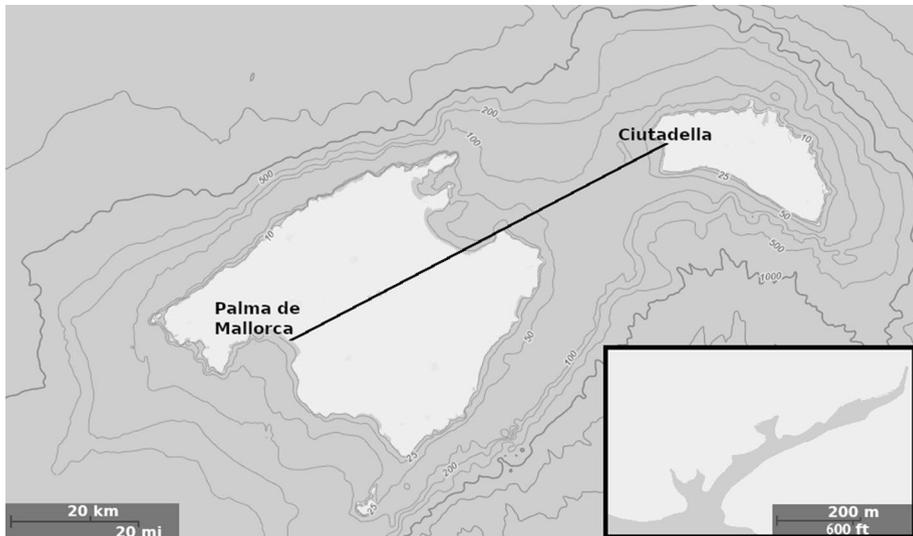


Fig. 1 Mallorca and Menorca (Balearic Islands). The regional atmospheric radiosounding station is located in Palma. The bottom-right inset is a close-up view of Ciutadella harbour, where the meteotsunamis occur. SW–NE oriented line indicates the optimal direction of propagation of precursor atmospheric gravity waves (see text). The bathymetry is indicated with labeled isolines in meters (GEBCO Compilation Group 2019)

This general meteotsunami generation scheme works indeed very well for the Ciutadella harbour context (see Fig. 1 for the particular geographical configuration). Several case studies (Ramis and Jansà 1983; Jansà 1986; Monserrat et al. 2006; Jansà et al. 2007; Ličer et al. 2017) show that rissaga events are simultaneous to a basic wind from the W-SW direction (the inlet orientation) presenting strong shear across the tropospheric column, and with a temperature vertical profile clearly shaping the aforementioned changes of stability. At synoptic scale, these flow characteristics require a cold upper-level trough to the west of the Balearic Islands accompanied or shortly preceded by weak circulation from the south at low levels. This southern circulation carries warm African air and generates a strong temperature inversion over the western Mediterranean sea (Ramis and Jansà 1983). On the other hand, the 55-km-long 80-m-deep Menorca channel, with a corresponding shallow-water wave speed of 28 m s^{-1} , favours the crucial action of the Proudman resonance mechanism. Additionally, the 1 km-long 5 m-deep harbour, with a fundamental eigenperiod of 10.5 min, guarantees strong coupling with the ocean waves impinging on the port with a similar frequency.

At present, there are a few operational prediction systems in place that focus on Ciutadella harbour rissagues. A helpful recapitulation can be found on Romero et al. (2019) which also presents the latest system added to the operational set. First, the national weather service of Spain (*Agencia Estatal de Meteorología*, AEMET) daily monitors the aforementioned synoptic circulation pattern prone to rissaga generation. The reliability of these warnings improves for the short term when clear signals of gravity wave activity can be confirmed on satellite images. However, quantitative estimation of the possible rissaga cannot be clearly established as it is more an educated guess of the forecaster on duty based on skill and experience. Second, a full 3D high-resolution atmosphere–ocean coupled model is applied in SOCIB/BRIFS (Balearic Islands Coastal Observing and Forecasting

System/Balearic RIssaga Forecasting System).² The SOCIB/BRIFS was described in detail and successfully tested for the June 2006 meteotsunami event by Renault et al. (2011). From a statistical point of view, the system reveals a good skill to recognize rissaga situations but a systematic tendency to underestimate the wave height (Romero et al. 2019). It is worth mentioning that such a system is very expensive computationally speaking, as the necessary grid resolutions require long simulation times and great amounts of dedicated memory. The third and last fully deployed approach,³ devised by these same authors, is the TRAM-rissaga system (Romero et al. 2019). This consists of a highly simplified atmosphere-ocean modeling system where shallow-water models over Menorca channel and Ciutadella harbour are forced with the atmospheric signal generated with a dry non-hydrostatic fully compressible numerical model. In fact, the TRAM system numerically reproduces the scheme of Šepić et al. (2015) by first generating the atmospheric gravity waves involved on meteotsunamis and then driving the propagation and amplification of long ocean waves until they resonate with the coastal inlet. For this application only the 2D version of the TRAM model is required as it focuses on the relevant Palma-Ciutadella vertical section (see Fig. 1). In addition, the system is initialized with a single observed or forecast thermodynamic sounding over the Balearic Islands. The use of a 2D atmospheric model coupled with the shallow-water equations was devised as a pragmatic and computationally fast approach to predict the occurrence and magnitude of rissagues. After testing, TRAM-rissaga shows valuable skill for the recognition of rissaga risk situations and their classification as weak, moderate or intense, at a much lower cost than the BRIFS system (see Romero et al. 2019 for details).

In this paper we develop a new strategy to predict rissagues on Ciutadella harbour that relies on the physical properties of the atmosphere that are clearly linked to meteotsunami generation. Once again, we used a radiosounding over the Balearic Islands to automatically identify, first of all, environmental characteristics prone or adverse to rissaga generation. Second, in order to produce an actual rissaga forecasting system we need a method able to link quantitatively these atmospheric ingredients (based on wind and temperature profiles) to the wave height recorded at Ciutadella harbour. An artificial neural network (NN to simplify notation) arises as the best suited algorithm to comply with these two requirements (to identify and to quantify), as neural networks are well known and praised for their pattern recognition abilities. It is also worth noticing that once properly trained a NN is an extremely computationally cheap method, facilitating its future mass application. Some examples of NN application in Meteorology are: Marzban and Stumpf (1996) that predicts tornadoes from Doppler radar signal attributes; Hall et al. (1999) predict the probability and intensity of precipitation in Dallas-Fort Worth (Texas) with mesoscale weather forecast model data and recent radiosounding data as inputs; Maqsood et al. (2004) that examines the use of an ensemble made of NNs to predict temperature, wind speed and relative humidity; Abhishek et al. (2012) that explores the applicability of this tool using different transfer functions, hidden layers and neurons to forecast maximum temperature for the 365 days of a year; and Nakajo et al. (2017) that tests the applicability of a NN to forecast meteotsunamis with observed atmospheric pressure data.

The paper has been structured as follows: A description of the new NN-based method used to predict meteotsunamis in Ciutadella is provided in Sect. 2. Main results in terms of

² See: <http://www.socib.es/index.php?seccion=modelling&facility=rissagaforecast>.

³ Operational rissaga forecasts by the TRAM system are available at <http://meteo.uib.es/rissaga>.

its skill before rissaga and non-rissaga situations are examined and discussed in Sect. 3. Finally, some conclusions and future lines of work are presented in Sect. 4.

2 Methodology: neural network set-up

An artificial neural network is a computing system capable of learning to execute a task as a biological NN would learn to: observing and trying to reproduce reality without a rulebook (Bishop et al. 1995; Haykin and Network 2004). The early goal of a NN was very broad: solving problems as a human brain would. Nowadays the goal has shifted to more specialized tasks such as modeling, time series prediction, data processing (filtering, clustering, etc.), robotics, computer numerical control and classification (pattern and sequence recognition, etc.). A NN is formed by at least three layers: input, output and hidden. The input layer is made of input neurons that represent the information that we want to feed to the network. The output layer contains the output neurons, that is the information that we want to predict with the network. Networks can have multiple hidden or middle layers. The hidden layer neurons represent a function that takes the output of all neurons in the previous layer and delivers a numeric value. This function is called activation function and introduces non-linear properties to our network. Neurons and these connections are associated with a weight that is adjusted following the learning process during the training phase. Signals travel from the input layer to the output layer after crossing the hidden layers multiple times during the training phase. This iteration cycle stops when the output produced by the neural network is close enough to the actual output according to a stopping criteria.

In this work we use a recurrent neural network which is good for processing sequential data aimed at predictions. A recurrent NN processes sequentially one input neuron at a time and it retains a memory of what has come previously in the sequence. Therefore, the network is able to learn the sequence's long-term dependencies (Ruhmelhart et al. 1986; Hochreiter and Schmidhuber 1997; Schmidhuber 2015). To be specific, we apply a NN using resilient backpropagation with weight backtracking method (RPROP+, Riedmiller and Rprop 1994). RPROP+ is a local adaptation learning scheme based on a gradient descent algorithm with a fast convergence rate. The scheme uses the sign of the gradients but not their magnitude to compute the weights updates. Depending on a heuristic function, the scheme can backtrack to a previous weight value. The backtracking can be done for all weights or just for some of them. To implement all NN computations we use the R statistics package Neuralnet (R Core Team 2018; Fritsch et al. 2019).

In order to successfully implement a NN to predict rissagues, we need to properly train the network. As previously stated, we want to use the data provided by radiosoundings to predict the potential harbour oscillation, therefore we need a database that contains past Palma radiosounding variables and the associated observed wave height at Ciutadella. For a previous study we compiled a suitable database using different sources (see Romero et al. 2019, Section 3.1 for all the details). It is worth mentioning that, in Ciutadella, a rissaga event refers to a seiche greater than 70 cm. Having this 70 cm threshold in mind, our database is made up of 126 rissaga days, extending with some gaps from July 1981 to July 2018 (plus one old case from September 1975) complemented with 549 non-rissaga days along the December 2016–July 2018 period. This database just includes those events for which at least one of the thermodynamic radiosoundings launched at 00 and 12 UTC from Palma station is available. Note that only one of the two daily soundings is used to characterize a database event, preferably the one closer in time to the moment of maximum daily

oscillation at Ciutadella. Also note that the seiche oscillation magnitude for some events of the 80s and 90s contains significant uncertainty as they were not directly measured, but simply estimated based on witnesses or newspapers news.

The input layer of our NN should contain the environmental key features behind rissaga generation. Recall these traits are basically the wind direction and vertical shear, and the vertical profile of static stability. Therefore, we used the horizontal wind components at different vertical levels given by the radiosounding as a proxy for the wind profile characteristics, whereas the stability is inferred from the temperature data at the different levels. This configuration does not take into account the atmospheric humidity so we refer to it as the dry scheme. However, it is known that in some rissaga events the humidity profile also plays a relevant role. Humidity is intrinsically linked to the development of moist convection, as large amounts of moisture at low levels are needed to build convective or latent instability in the tropospheric column. Under these situations, the surface pressure variations are not only related to gravity wave activity but also to pressure jumps imposed by propagating convective systems, including small squall lines. Typically, these convective systems are triggered by greatly amplified gravity waves. For instance, Jansá et al. (2007) associated the extraordinary character of the 15 June 2006 meteotsunami event to the north-eastward propagation of a mesoscale convective system (MCS) over the Balearic Islands. Meteotsunamis triggered or amplified by MCS have also been documented for other locations (e.g. Wertman et al. 2014; Bailey et al. 2014). Consequently, in this work we also considered a second NN configuration, called the wet scheme, in which radiosounding-derived dew point depression⁴ data was ingested in addition to winds and temperatures.

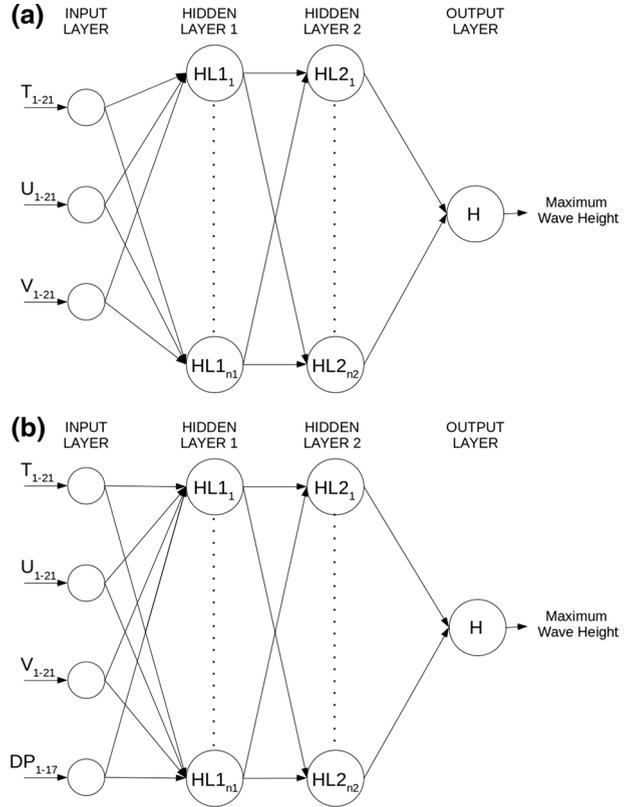
In summary, we devised two different approaches: (1) a dry scheme where the NN input layer consists of the radiosounding temperature, u-wind (zonal) and v-wind (meridional) components vertically interpolated to a fixed set of 21 pressure levels,⁵ the same set used for the operational outputs of the NCEP/Global Forecasting System (GFS) model; and (2) a wet scheme where the dry NN input layer was completed with the dew point depressions throughout the 1000–300 hPa range (see Fig. 2 for a graphical representation of both NNs schemes). We excluded the stratospheric levels—humidity is very low at high altitudes—because they do not add relevant information to the NN while they could hinder its learning process. In conclusion, we have a 63-neuron input layer for the dry scheme and a 80-neuron input layer for the wet scheme. Both schemes have a 1-neuron output layer that corresponds to the maximum wave height (crest-to-trough difference in a \approx 5-min interval) measured at Ciutadella harbour during the whole day.

In fact, we designed a NN for each scheme, dry and wet, with the same input and output layers, but including 2-hidden layers with an undetermined number of neurons (see Fig. 2). Only two hidden layers were chosen to maintain the NN design as simple as possible. On each hidden layer the following range of number of neurons was tested: (1) on hidden layer 1 from 100 to 350 by 50 and (2) on hidden layer 2 from 10 to 100 by 10. The training stopping criteria—the threshold for the partial derivatives of the error function—was also determined during the design of the NN as it can affect the ability of the model to predict outside the range of the training dataset. Specifically, these two training stopping criteria

⁴ The dew point depression is the difference between the actual and dew point temperatures at a certain height in the atmosphere, thus a measure of the closeness of the air parcel to saturation.

⁵ Namely, these pressure levels are: 1000, 975, 950, 925, 900, 850, 800, 750, 700, 650, 600, 550, 500, 450, 400, 350, 300, 250, 200, 150 and 100 hPa.

Fig. 2 Graphical representation of both NNs configurations: **a** dry scheme and **b** wet scheme



values were tested: 0.005 and 0.001. The logistic function⁶ was always used as activation function. Wave heights and all input variables were normalized, linearly scaling the data into the 0–1 interval, suitable for our activation function. Taking as reference the minimum and maximum recorded values of every input and output variable, expanded by 20% to ensure applicability outside the range used for training and testing, we assigned 0–80% of the minimum, and 1–120% of the maximum. The whole database, 126 rissaga days plus 549 non-rissaga days, was sliced into two sets: a training database (70% of the cases) and a testing database (the remaining 30%). This splitting of the original database into two subsets was done randomly, but only once, so we always included the same events in each subset. We checked that the R package sample function used for this splitting has effectively separated the database into two subsets that correctly represent the characteristics of the whole database (see Fig. 3). Each NN was trained with the training subset and then validated with the testing subset. For each scheme, we chose the best NN configuration based on the following criteria during the validation process (see Fig. 4): (1) limited spread: we want the NN prediction as close to the observation as possible, and therefore, we want to minimize the number of events where the difference between observation and prediction in

⁶ A logistic function is a mathematical function that has a characteristic S-shaped curve. The logistic function was first introduced in Pierre-François (1838).

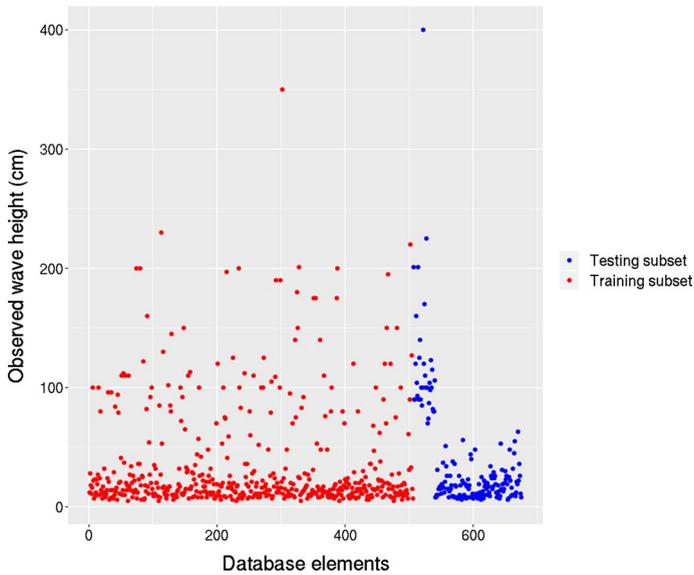


Fig. 3 Observed wave height (in cm) for the whole database, rissaga (> 70 cm) and non-rissaga (< 70 cm) events. The training subset is plotted with red dots and the testing subset with blue dots

Fig. 4 Flow chart of the selection process of the best NN configuration

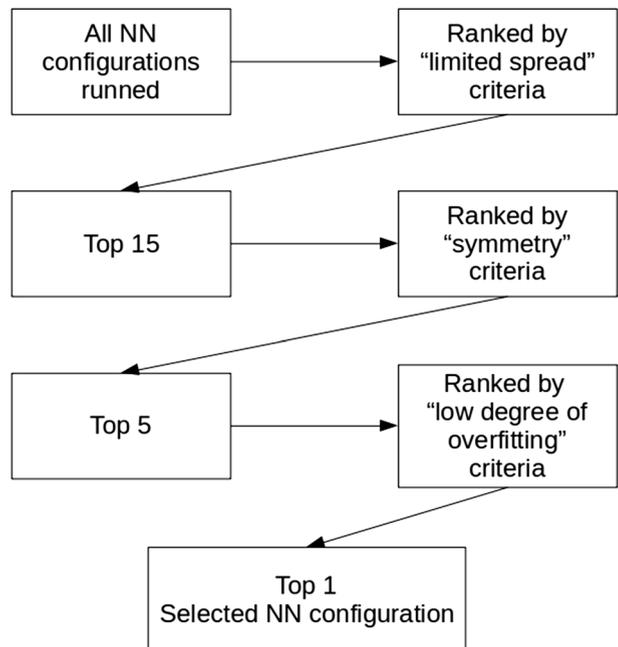


Table 1 Performance statistics for the dry and wet NN schemes and the TRAM-rissaga system: the observed mean height (OBS mean), the simulated mean height (SIM mean), the root-mean-square error (RMSE) and the linear correlation coefficient (CC)

	OBS mean (cm)	SIM mean (cm)	RMSE (cm)	CC
Dry NN	39.2	36.7	44.2	0.307
Wet NN	39.2	33.4	44.3	0.315
TRAM-rissaga	37.0	42.8	45.4	0.431

absolute value is more than 50 cm,⁷ (2) symmetry: we want a forecasting system as unbiased as possible, therefore we seek to have as few over-predictions as under-predictions, and (3) low degree of overfitting: we want a NN capable of generalizing from our training subset, therefore we discard the configurations that fit the training subset too closely (i.e. a spread lower than 10 cm, a subjective choice). Combining these criteria, we were able to determine semi-objectively the optimal configuration of the NN for each scheme.

3 Results and discussion

After testing several NN configurations and applying the criteria described in Sect. 2, we kept the following structure:

- Dry scheme: a 63-neuron input layer, a 100-neuron and a 10-neuron hidden layers, and a 1-neuron output layer. The training stopping criteria is 0.005.
- Wet scheme: a 80-neuron input layer, a 350-neuron and a 50-neuron hidden layers, and a 1-neuron output layer. The training stopping criteria is 0.001.

Since we have the TRAM-rissaga system results available for comparison from Romero et al. (2019), we add them to our discussion. It is worth noticing that for the TRAM-rissaga system the whole database was available for testing, unlike for the NN framework, where we necessarily had to split the database into two subsets: training and testing. This contrast between methodologies can affect the results discussion since we have a smaller population to test the performance of the NNs. Having a small sample size can influence the results of most verification indices. However, this is an inescapable limitation because we are constrained by the only existing rissaga database and by the need to train the networks.

Considering some basic statistics, both NN schemes and the TRAM-rissaga system perform fairly well (see Fig. 5 and Table 1). The difference between observed and simulated mean height at Ciutadella harbour is less than 6 cm for all three systems, although both NN schemes slightly underpredict on average, while TRAM-rissaga overpredicts. The root-mean-square error lies below 50 cm for the three systems, being the TRAM-rissaga value the biggest by 1 cm. All the obtained linear correlation coefficients indicate a limited association between the observed and simulated populations; the TRAM-rissaga coefficient value is a bit higher than both NN schemes values, but the latter still indicate a weak to moderate association. It is worth mentioning that all three methods severely underpredict the 400-cm wave height of

⁷ This is a subjective error threshold, also used in Romero et al. (2019).

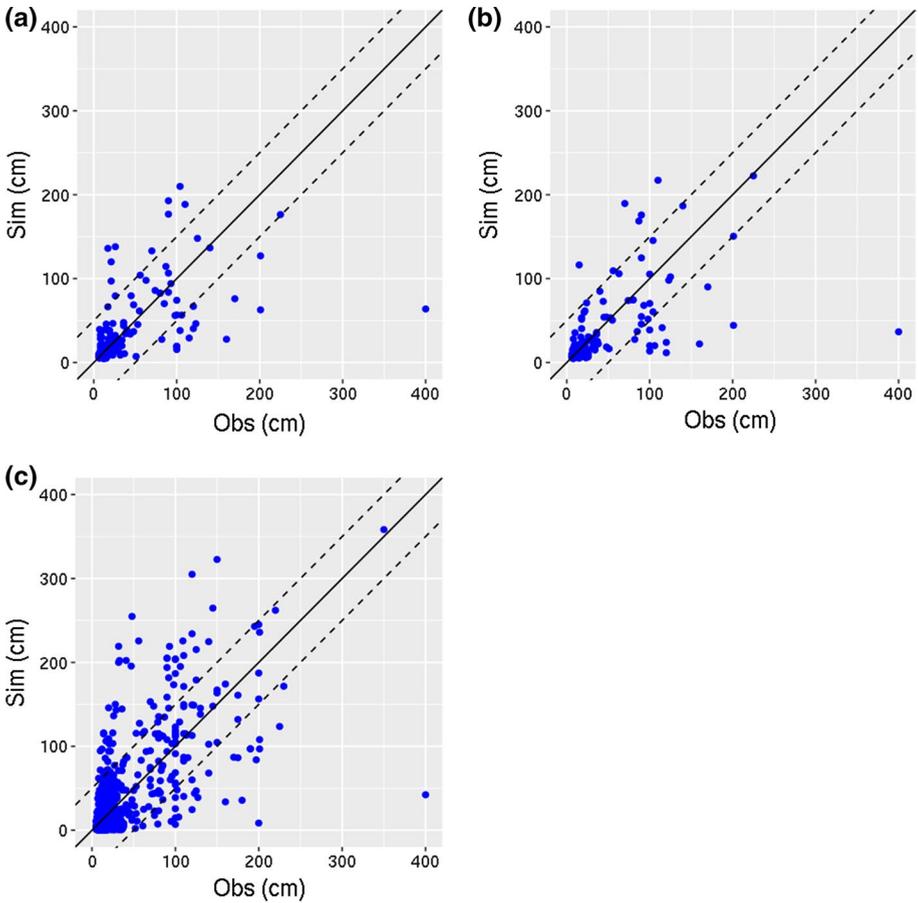


Fig. 5 **a** Dry NN scheme, **b** wet NN scheme and **c** TRAM-rissaga system simulated water oscillation against observed water oscillation for each corresponding testing database. Dashed lines next to the diagonal (solid line) encompass the ± 50 cm limits

Table 2 Two different scale categories for the wave height values: dichotomous and expanded

Scale	Category	Wave height (cm)
Dichotomous	Non-rissaga	< 70
	Rissaga	≥ 70
Expanded	Small oscillations	< 20
	Moderate oscillations	$\in [20, 70)$
	Ordinary rissaga	$\in [70, 100)$
	Intense rissaga	$\in [100, 200)$
	Extreme rissaga	≥ 200

Table 3 Contingency table for observed event and forecast event in a 2 × 2 problem

	Observed		
	Yes	No	Total
Forecast			
Yes	Hit	False alarm	Forecast yes
No	Miss	Correct rejection	Forecast no
Total	Observed yes	Observed no	Total

the 15 June 2006 record event. In particular, the NN dry scheme predicts 64 cm, the NN wet scheme 37 cm and TRAM-rissaga 42 cm, so a difference with the observation greater than 300 cm on all three methods (see Fig. 5). A possible explanation of this deficiency is that this record rissaga event was strongly linked to a mesoscale convective system crossing the Balearic Islands (Jansá et al. 2007). Therefore the NN dry scheme and the TRAM-rissaga are, by definition, not suitable to properly reproduce this kind of events. However, the NN wet scheme does include a certain representation of moist processes so the reasons it does not better capture this event are not properly known. As we will discuss later, a plausible explanation is that the number convectively driven events is so small in the database that the NN is not properly trained to successfully capture them.

A different approach for evaluating the NN forecasts is to use verification scores and compare their values with the TRAM-rissaga system results. Figure 6 displays several complementary verification scores (see Jolliffe and Stephenson 2012 for a comprehensive guide on forecast verification) for both NN schemes and the TRAM-rissaga system, and for two different rissaga categorization scales (see Tables 2, 3). This choice of two categorization scales is motivated by the following: (1) a dichotomous scale of event and non-event naturally arises to test the potential of a method to discriminate rissaga situations from ordinary days; (2) an expanded scale is required when we want to check the ability of the method to predict the intensity of the event. The five rissaga classes of the expanded scale (wave height intervals) are chosen according to the expected risk and possible damages to the port and its activities. In fact, all current operational systems use the same scale for issuing specific warnings to Ciutadella harbour. It is worth noticing that both scales, dichotomous and expanded, are unrelated. In other words, we either analyse the rissaga discrimination power of the forecasting system or its ability to provide quantitative details.

$$ACC = \frac{\text{Hits} + \text{Correct Negatives}}{\text{Total}} \tag{1}$$

$$BIAS = \frac{\text{Hits} + \text{False Alarms}}{\text{Hits} + \text{Misses}} \tag{2}$$

$$POD = \frac{\text{Hits}}{\text{Hits} + \text{Misses}} \tag{3}$$

$$FAR = \frac{\text{False Alarms}}{\text{Hits} + \text{False Alarms}} \tag{4}$$

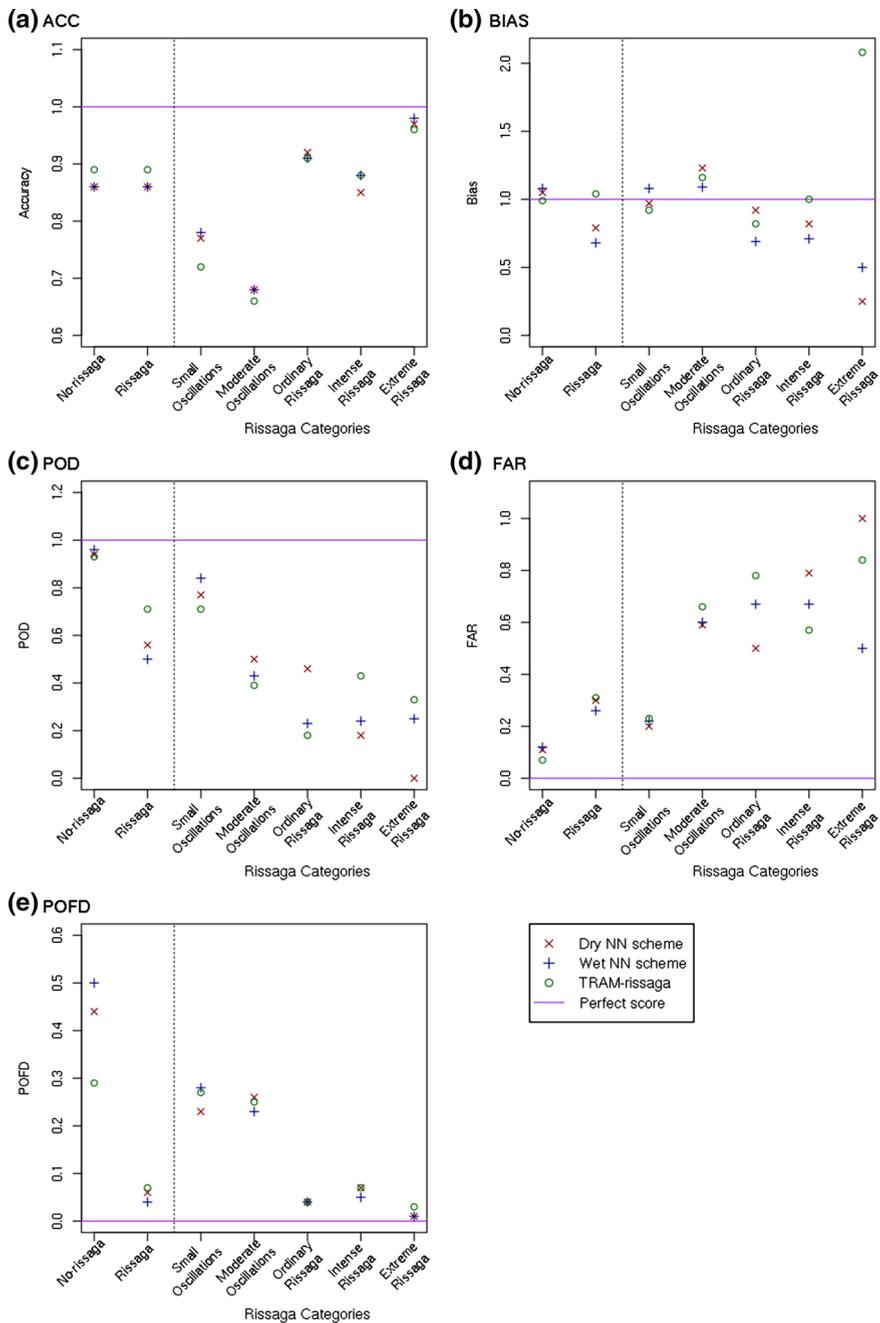


Fig. 6 **a** Accuracy, **b** bias, **c** probability of detection, **d** false alarm ratio and **e** probability of false detection for the dry and wet NN schemes and the TRAM-rissaga system (red crosses, blue pluses and green circles, respectively). The purple solid line indicates the perfect score for each corresponding verification index. Results for the dichotomous categorization scale are plotted to the left of the vertical dotted line and those for the expanded categorization scale on the right side (see Table 2 for more details on these scales). Note the different value ranges encompassed by the vertical axes

$$\text{POFD} = \frac{\text{False Alarms}}{\text{Correct Negatives} + \text{False Alarms}} \quad (5)$$

The accuracy (ACC, Fig. 6a, Eq. 1) indicates what fraction of the forecasts are correct, so a perfect score would be 100%. All three methods show good skill in discriminating between non-rissaga and rissaga events, in fact 85% of the non-rissaga and the rissaga forecasts were correct for both NN schemes and almost 90% for the TRAM-rissaga method. Overall, we obtain that more than 65% of the forecasts were correct for each expanded category, being the ordinary, intense and extreme rissagues the best forecast categories, with at least 85% of correct forecasts. The exceptionally good result of the three methods for the extreme rissaga category is influenced by the small number of events that fall into this category.

The bias (BIAS, Fig. 6b, Eq. 2) reveals whether a forecast system has a tendency to underforecast (bias value below 1) or overforecast (bias value over 1) events. All three methods are very well calibrated for the non-rissaga category. The TRAM-rissaga maintains this good calibration for the rissaga category, although both NN schemes underforecast this class. Looking at the expanded scale categories, the bias results reveal good skill for all categories except for the extreme rissaga, that is again influenced by having a small number of events in this extraordinary class. In addition, it is worth mentioning that the TRAM-rissaga method overforecasts while both NN schemes underforecast in this category.

The probability of detection (POD, Fig. 6c, Eq. 3) indicates the fraction of events that are correctly forecast, this index would reach a 100% for a perfect system. The POD value for the non-rissaga category is very good for all three methods, more than 90%. For the rissaga category, there is greater spread among methods. The best skill corresponds to TRAM-rissaga with a POD value around 75% while both NN schemes lie around 55%. The expanded scale categories show a diversity of score values among the forecasting strategies. By far all three methods show the best skill for the small oscillations category, with a POD value over 70%, while the rest of categories contain POD values below 60% and progressively declining as the magnitude of the rissaga is increased.

The false alarm ratio (FAR, Fig. 6d, Eq. 4) measures the fraction of predicted events that did not occur, so ideally FAR values would lie close to 0%. All three systems are very well adapted at not producing false alarms in the non-rissaga category, the FAR values are below 10% for all three. On the rissaga category, all methods exhibit very similar FAR values of approximately 30%. Like in the POD, on the expanded scale all of them show the best skill in the small oscillations category, with a FAR value below 25%. For the rest of categories FAR values situate over 40%, increasing spread and magnitude as the wave height augments. In particular, FAR scores for the extreme rissaga category covers a range of 50%, with the NN wet scheme having the best FAR index (approximately 50%) and the TRAM-rissaga the worst one (almost 100%). Again, this result is heavily influenced by the small number of events that fall into this category.

Finally, the probability of false detection (POFD, Fig. 6e, Eq. 5) refers to the fraction of observed non-events that are incorrectly forecast as events; again the POFD scores should ideally remain close to 0%. The non-rissaga category displays POFD values for all three prognostic methods quite far from 0%. In fact the best behavior is exhibited by the TRAM-rissaga system, with a score of approximately 30%, while the POFD values for both versions of the NN are around 45–50%. In contrast, the rissaga category presents very good results for any forecasting system, with POFD values below 10%. Focusing on the expanded scale categories, all methods perform very similarly for the five wave height

classes. The worst POFD scores, with values between 20 and 30%, are exhibited for the small and moderate oscillation classes. In contrast, the three rissaga categories present POFD values below 10%.

4 Conclusions and further work

This paper explores the potential of using a NN to forecast the meteotsunamis that occur in Ciutadella harbour as an interesting complement to traditional forecasting methods that rely on physical numerical simulations. A NN, once trained, becomes a very cheap method in terms of its computational needs, so a forecast is ready almost instantaneously once the input information is available. Such a prompt forecast, if skilled enough, is crucial to warn the port authorities with sufficient time as to permit the deployment of all kinds of preventive measures. Our results show that the NN not only provides a fast forecast but also a skilled one, at least comparable to the outcomes of other available approaches that are computationally more expensive. The results show that our former TRAM-rissaga atmosphere-ocean numerical approach is somewhat more accurate than both tested NNs. However, this difference in skill is small and for some rissaga categories the NNs even exhibited better verification scores for some forecasting attributes. The difference in performance could be very well compensated by having a forecast more promptly, enhancing its usefulness. Nevertheless, such a cost-benefit analysis should be properly done once the NN approach presented here has been running operationally and in conjunction with the other available techniques and the port authorities.

According to the examined verification scores, rather small differences between the performance of both NN schemes, dry and wet, can be inferred. The addition of humidity profiles in the input layer is in principle a physically sound strategy to enlarge the represented spectrum of rissaga generation mechanisms. But in practice, the improvement of the wet scheme over the dry scheme as revealed by the verification results is rather limited. A remarkable exception is obtained for the extreme rissaga category (see Fig. 6). This could be expected, since it is precisely for extraordinary events when the role of deep moist convection has been revealed the most crucial. It appears that the weak overall influence of the moisture factor is strongly tied to the available rissaga database. The database does not include enough events linked to convection-related generation or enhancement mechanisms as to supply a sufficient sample of cases during the NN training phase. The testing phase is even more scarcely populated by this type of events. Widening the database with older cases is a very difficult task due to the lack of radiosoundings and/or wave height records. However, we could use the whole database for training and perform the validation in real time as future events occur. With this approach, we would expect a considerable improvement in skill for both schemes and more noticeable benefits of the wet scheme over the dry scheme.

Having these results and conclusions in mind, the next step is to implement operationally an automatic prediction system for Ciutadella harbour based on both NNs. We propose to use as input data pseudo-radiosoundings extracted from GFS forecasts at the same location of Palma; that is, GFS-predicted vertical profiles of wind vector, temperature and dew point depression at the same pressure levels used for the training of the NNs. In addition, fast and sequentially updated wave height predictions could be produced by processing the latest GFS forecasts available over consecutive time steps (e.g. at hourly intervals). These lagged predictions could be combined to generate an ensemble forecast of harbour

oscillation amplitude that can be exploited to infer the rissaga risk in probabilistic terms, or as function of the period of the day. The reasons for this strategy are twofold: (1) to better cope with the inherent uncertainties of having an appropriate representative sounding; and (2) to adapt the meteotsunami forecasting system to rapidly evolving synoptic situations.

Acknowledgements We thank our department colleague Dr. José Luis Rosselló who provided insight into neural networks. Two Spanish institutions, Agencia Estatal de Meteorología (AEMET) and Ports de Balears, are acknowledged for helping to compile the database of rissaga events. Bathymetric data was originally plotted with the NCEI/NOAA Bathymetric Data Viewer (Fig. 1) at <https://maps.ngdc.noaa.gov/viewers/bathymetry>. This research was sponsored by PCIN-2015-221 (METEOforSIM) and CGL2017-82868-R (COAST-EPS) projects, both of them funded by the Spanish *Ministerio de Economía, Industria y Competitividad* and partially supported with FEDER, Spain funds.

References

- Abhishek K, Singh M, Ghosh S, Anand A (2012) Weather forecasting model using artificial neural network. *Procedia Technol* 4:311–318
- Bailey KE, DiVeglio C, Welty A (2014) An examination of the June 2013 east coast meteotsunami captured by NOAA observing systems. Executive Summary
- Bishop CM et al (1995) Neural networks for pattern recognition. Oxford University Press, Oxford
- Fritsch S, Guenther F, Wright MN (2019) neuralnet: training of neural networks. R package version 1.44.2. <https://CRAN.R-project.org/package=neuralnet>. Accessed 5 Nov 2019
- GEBCO Compilation Group (2019) GEBCO 2019 grid. <https://doi.org/10.5285/836f016a-33be-6ddc-e053-6c86abc0788e>. Accessed 18 Nov 2019
- Hall T, Brooks HE, Doswell CA III (1999) Precipitation forecasting using a neural network. *Weather Forecast* 14(3):338–345
- Haykin S, Network N (2004) A comprehensive foundation. *Neural Netw* 2(2004):41
- Hochreiter S, Schmidhuber J (1997) Long short-term memory. *Neural Comput* 9(8):1735–1780
- Horvath K, Šepić J, Prtenjak MT (2019) Atmospheric forcing conducive for the Adriatic 25 June 2014 meteotsunami event. In: *Meteorology and climatology of the Mediterranean and Black Seas*. Springer, pp 97–117
- Jansá A (1986) Marine response to mesoscale-meteorological disturbances: the June 21, 1984, event in Ciutadella (Menorca). *Rev Meteorol* 7:5–29
- Jansá A, Monserrat S, Gomis D (2007) The rissaga of 15 June 2006 in Ciutadella (Menorca), a meteorological tsunami. *Adv Geosci* 12:1–4
- Jolliffe IT, Stephenson DB (2012) *Forecast verification: a practitioner's guide in atmospheric science*. Wiley, Hoboken
- Ličer M, Murre B, Troupin C, Krietemeyer A, Jansá A, Tintoré J (2017) Numerical study of balearic meteotsunami generation and propagation under synthetic gravity wave forcing. *Ocean Model* 111:38–45
- Maqsood I, Khan MR, Abraham A (2004) An ensemble of neural networks for weather forecasting. *Neural Comput Appl* 13(2):112–122
- Marzban C, Stumpf GJ (1996) A neural network for tornado prediction based on doppler radar-derived attributes. *J Appl Meteorol* 35(5):617–626
- Monserrat S, Rabinovich A et al (2006) Meteotsunamis: atmospherically induced destructive ocean waves in the tsunami frequency band. *Nat Hazards Earth Syst Sci* 6(6):1035–1051
- Nakajo S, Yamaguchi R, Kim S, Tsujimoto G (2017) Basic study on real-time prediction of meteotsunami using artificial neural network model calibrated by stationary measurement data. In: *Asian and Pacific Coast 2017-proceedings of the 9th international conference on Apac 2017*. World Scientific, pp 292–301
- Pierre-François V (1838) Notice sur la loi que la population poursuit dans son accroissement. *Corresp Math Phys* 10:113–121
- R Core Team (2018) R: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
- Rabinovich AB (2010) Seiches and harbor oscillations. In: *Handbook of coastal and ocean engineering*. World Scientific, pp 193–236

- Rabinovich AB, Monserrat S (1998) Generation of meteorological tsunamis (large amplitude seiches) near the balearic and kuril islands. *Nat Hazards* 18(1):27–55
- Ramis C, Jansà A (1983) Condiciones meteorológicas simultáneas a la aparición de oscilaciones del nivel del mar de amplitud extraordinaria en el mediterráneo occidental. *Rev Geofis* 39:35–42
- Renault L, Vizoso G, Jansà A, Wilkin J, Tintoré J (2011) Toward the predictability of meteotsunamis in the balearic sea using regional nested atmosphere and ocean models. *Geophys Res Lett* 38(10):L10601
- Riedmiller M, Rprop I (1994) Rprop description and implementation details. Technical report
- Romero R, Vich M, Ramis C (2019) A pragmatic approach for the numerical prediction of meteotsunamis in Ciutadella harbour (Balearic Islands). *Ocean Model* 142:101441
- Ruhmelhart D, Hinton G, Williams R (1986) Learning representations by back-propagation errors. *Nature* 323:533–536
- Schmidhuber J (2015) Deep learning in neural networks: an overview. *Neural Netw* 61:85–117
- Šepić J, Vilibić I, Rabinovich AB, Monserrat S (2015) Widespread tsunami-like waves of 23–27 june in the mediterranean and black seas generated by high-altitude atmospheric forcing. *Sci Rep* 5:11682
- Tintoré J, Gomis D, Alonso S, Wang DP (1988) A theoretical study of large sea level oscillations in the western mediterranean. *J Geophys Res Oceans* 93(C9):10797–10803
- Vilibić I, Šepić J, Rabinovich AB, Monserrat S (2016) Modern approaches in meteotsunami research and early warning. *Front Mar Sci* 3:57. <https://doi.org/10.3389/fmars.2016.00057>
- Wertman CA, Yablonsky RM, Shen Y, Merrill J, Kincaid CR, Pockalny RA (2014) Mesoscale convective system surface pressure anomalies responsible for meteotsunamis along the us east coast on June 13th, 2013. *Sci Rep* 4:7143

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.